Low-Entanglement Remote State Preparation

Igor Devetak* and Toby Berger
Department of Electrical and Computer Engineering
Cornell University, Ithaca, New York 14850

February 27, 2001

Abstract

The exact low-entanglement remote state preparation (RSP) ebits vs. bits tradeoff curve [1] is found using techniques of classical information theory. This further gives rise to a lower bound for the high-entanglement region.

The problem addressed here is that of remote preparation of a qubit state where the sender (a.k.a. Alice) posesses complete classical knowledge of the qubit state to be prepared in Bob's laboratory [1]. We have two types of resources at our disposal: entanglement (ebits) between Alice and Bob and classical bits of forward communication from Alice to Bob. In addition, it has been assumed that Bob is allowed classical communication to Alice, since such backward classical communication is unhelpful for teleportation. In the context dealt with here backward communication is also of no use.

One may ask about the optimal tradeoff curve in the ebits vs. bits plane necessary for Bob to be able to produce a copy of the state which is perfect in the asymptotic sense. One point on this graph is one ebit vs. two bits which corresponds to teleportation [2]. Lo [3] notes that even when Alice has complete classical knowledge of the state to be teleported, she cannot get away with $2-\epsilon$ bits of forward communication. Otherwise, by having used the qubit teleported to encode 2 classical bits via dense coding [4], she would effectively be able to convey ϵ classical bits by virtue of mere entanglement, thus violating causality.

We concentrate on the low entanglement region where the number of ebits is less than one. The method used by Bennett et al.[1] is to send partial classical information about the qubit state, thus reducing its posterior von Neumann entropy (i.e. the von Neumann entropy of the posterior density operator reflecting Bob's knowledge about the qubit state having recieved the classical information). This may be used to Schumacher compress a block of such qubits [5], and teleport them at a rate corresponding to the posterior von Neumann entropy. The coding scheme they use is suboptimal, and, as we will show, may be greatly improved on by employing a scheme related to classical rate-distortion coding [6] [7]. We show, moreover, that our result cannot be improved on.

We now formulate the coding problem. The source is described by a random vector $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ and we take the individual X_i to be independent and identically distributed (i.i.d.), each taking values x in the set $\mathcal{X}$ with probability density $p(x)$. Thus the probability density distribution for $\mathbf{X}$ is $p(\mathbf{x}) = \prod_i p(x_i)$. In our case $\mathcal{X}$ is the surface of a sphere parametrized by spherical polar coordinates $(\theta_x, \phi_x) \in [0, \pi] \times [0, 2\pi]$, and

*Electronic address: igor@ece.cornell.edu

$p(x) = \frac{1}{4\pi}$ corresponds to the uniform distribution. For convenience we refer to the north pole $(\theta_x, \phi_x) = (0, 0)$ as $x = 0$. In the quantum interpretation $\mathcal{X}$ corresponds to the Hilbert space of a qubit, and each $x \in \mathcal{X}$ is a Bloch sphere representation of a pure qubit state $|x\rangle = \sqrt{\frac{1+\cos\theta_x}{2}}|0\rangle + e^{i\phi_x}\sqrt{\frac{1-\cos\theta_x}{2}}|1\rangle$.

Elements $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ of $\mathcal{X}^n$ are called *source words* of length n, and the x_i are called *letters*. We map the source $\mathbf{X}$ onto a set $B_n = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_K\}$, called a *source code* of *size* K and *blocklength* n, of reproducing *codewords*. The *rate* of the code is defined as $R = n^{-1}\log_2 K$, and it signifies the number of bits per source letter needed to specify the reproducing codeword. Each source word $\mathbf{x}$ gets mapped into a unique $\mathbf{y} \in B_n$ in such a way that the posterior von Neumann entropy of the source.

$$\overline{S}(B_n) = \frac{1}{n} E_{\mathbf{Y}} S(E_{\mathbf{X}|\mathbf{Y}}|\mathbf{X}\rangle\langle\mathbf{X}|) \tag{1}$$

is minimized. Here $\mathbf{Y}$ is the random variable associated with the probability distribution on the set of codewords B_n induced by our map. $E_{\mathbf{Y}}$ denotes the expectation value over the random vector $\mathbf{Y}$, and $E_{\mathbf{X}|\mathbf{Y}}$ is the conditional expectation over $\mathbf{X}$ given the value of $\mathbf{Y}$. A rate-entropy pair (R, S) is *achievable* iff there exists a sequence of source codes B_n of rate R and increasing blocklength n such that

$$\lim_{n\to\infty} \overline{S}(B_n) \leq S \tag{2}$$

Then the *rate-entropy function* $R(S)$ is defined as the infimum of all R for which (R, S) is achievable. Using Schumacher compression and teleportation this implies the achievability of the curve $(S, R(S)+2S)$, $S \in (0, 1]$ in the ebits vs. bits plane. Conversely, by the optimality of Shumacher compression and Lo's dense coding/causality argument, this ebit vs. bit tradeoff cannot be improved on. Our goal is, therefore, to identify $R(S)$.

We consider the following problem: For a given blocklength n and probability density $p(\mathbf{x})$ of $\mathbf{X}$ defined above find

$$R_n(S) = \frac{1}{n} \inf_{Q(\mathbf{y}|\mathbf{x}):S(Q)=nS} I(Q) \tag{3}$$

where $I(Q)$ is the Shannon mutual information

$$I(Q) = \iint d\mathbf{x}d\mathbf{y}p(\mathbf{x})Q(\mathbf{y}|\mathbf{x})\log\frac{Q(\mathbf{y}|\mathbf{x})}{q(\mathbf{y})} = \iint d\mathbf{x}d\mathbf{y}q(\mathbf{y})P(\mathbf{x}|\mathbf{y})\log\frac{P(\mathbf{x}|\mathbf{y})}{p(\mathbf{x})} \tag{4}$$

and

$$S(Q) = \int d\mathbf{y}q(\mathbf{y})S\left(\int d\mathbf{x}P(\mathbf{x}|\mathbf{y})|\mathbf{x}\rangle\langle\mathbf{x}|\right) \tag{5}$$

The probability density for the marginal $\mathbf{Y}$ distribution is given by $q(\mathbf{y}) = \int d\mathbf{x}p(\mathbf{x})Q(\mathbf{y}|\mathbf{x})$ and the conditional distribution for $\mathbf{X}$ given $\mathbf{Y}$ is $P(\mathbf{x}|\mathbf{y}) = p(\mathbf{x})Q(\mathbf{y}|\mathbf{x})/q(\mathbf{y})$.

Using $p(\mathbf{x}) = (4\pi)^{-n}$, changing the integration variables from $(\mathbf{x}, \mathbf{y})$ to $(\mathbf{\Omega}, \mathbf{y})$, where $\mathbf{\Omega} \in \mathcal{R}^n$, $\mathcal{R}$ being the group of rotations in $\mathcal{X}$, for which $\mathbf{x} = \mathbf{\Omega}(\mathbf{y})$, and using the fact that von Neumann entropy is invariant under such rotations, the problem reduces to the Lagrange multiplier minimization of

$$F \equiv \int d\mathbf{y}q(\mathbf{y})\left(\mathcal{J}[P_{\mathbf{y}}] + \lambda\mathcal{S}[P_{\mathbf{y}}]\right) \tag{6}$$

where

$$\mathcal{J}[P_{\mathbf{y}}] \equiv \int d\mathbf{\Omega} P_{\mathbf{y}}(\mathbf{\Omega}) \log P_{\mathbf{y}}(\mathbf{\Omega}) \tag{7}$$

$$\mathcal{S}[P_{\mathbf{y}}] \equiv S\left(\int d\mathbf{\Omega} P_{\mathbf{y}}(\mathbf{\Omega}) |\mathbf{\Omega}(\mathbf{y})\rangle\langle\mathbf{\Omega}(\mathbf{y})|\right) = S\left(\int d\mathbf{\Omega} P_{\mathbf{y}}(\mathbf{\Omega}) |\mathbf{\Omega}(\mathbf{0})\rangle\langle\mathbf{\Omega}(\mathbf{0})|\right) \tag{8}$$

and $P_{\mathbf{y}}(\mathbf{\Omega}) \equiv P(\mathbf{\Omega}(\mathbf{y})|\mathbf{y})$, under the constraint

$$\int d\mathbf{y} q(\mathbf{y}) P_{\mathbf{y}}(\Omega_{\mathbf{y}\to\mathbf{x}}) = \frac{1}{4\pi}, \ \forall \mathbf{x} \tag{9}$$

Here $\Omega_{\mathbf{y}\to\mathbf{x}}$ is the unique element of $\mathcal{R}^n$ that maps $\mathbf{y}$ onto $\mathbf{x}$. We now note that for any $P_{\mathbf{y}}(\mathbf{\Omega})$ there is a $\mathbf{y}$-independent distribution $P_{\mathbf{y_0}}(\mathbf{\Omega})$ where $\mathbf{y} = \mathbf{y_0}$ minimizes $\mathcal{J}[P_{\mathbf{y}}] + \lambda\mathcal{S}[P_{\mathbf{y}}]$ that trivially satisfies (9) and by definition yields a value of F lower than or equal to $P_{\mathbf{y}}(\mathbf{\Omega})$. Therfore the minimum is achieved for $P_{\mathbf{y}}(\mathbf{\Omega})$ independent of $\mathbf{y}$, and we henceforth drop the $\mathbf{y}$ subscript from it. A corollary is that $q(\mathbf{y}) = \frac{1}{4\pi}$.

Parametrizing $\mathbf{\Omega}$ by $(\mathbf{u}, \mathbf{\Phi}) \in [-1,1]^n \times [0, 2\pi]^n$ with $u_i = \cos\theta_i$, we will further show that the optimum distribution $P_{\mathbf{\Phi}}(\mathbf{u}) \equiv P(\mathbf{u}, \mathbf{\Phi})$ has no $\mathbf{\Phi}$ dependence. We define

$$I(\mathbf{\Phi}) = \int d\mathbf{u} P_{\mathbf{\Phi}}(\mathbf{u}) \log P_{\mathbf{\Phi}}(\mathbf{u}) \tag{10}$$

and

$$\rho(\mathbf{\Phi}) = \int d\mathbf{u} P_{\mathbf{\Phi}}(\mathbf{u}) |\mathbf{u}, \mathbf{\Phi}\rangle\langle\mathbf{u}, \mathbf{\Phi}| \tag{11}$$

Our problem is equivalent to minimizing

$$S = S\left(\int d\mathbf{\Phi} \rho(\mathbf{\Phi})\right) \geq \int d\mathbf{\Phi} S(\rho(\mathbf{\Phi})) \tag{12}$$

while keeping $I = \int d\mathbf{\Phi} I(\mathbf{\Phi})$ fixed. The inequality in (12) is a consequence of the concavity of entropy. Given a distribution $P_{\mathbf{\Phi}}(\mathbf{u})$ we define $\mathcal{C}$ to be the convex hull of the set of all points $(I(\mathbf{\Phi}), S(\rho(\mathbf{\Phi})))$. Then (I, S) lies in or above $\mathcal{C}$, by (12). Consequently, there exist $\mathbf{\Phi_1}, \mathbf{\Phi_2}$ and $\alpha \in [0,1]$ such that the $\mathbf{\Phi}$-independent distribution defined by $P(\mathbf{u}) \equiv \alpha P_{\mathbf{\Phi_1}}(\mathbf{u}) + (1-\alpha) P_{\mathbf{\Phi_2}}(\mathbf{u})$ has the same I but lower or equal S as the original distribution.

We therefore restrict attention to $\mathbf{\Phi}$-independent distributions $P(\mathbf{u})$, and work in the basis $\{|0\rangle, |1\rangle\}^n$, where $|0\rangle$ corresponds to $u = 1$ and $|1\rangle$ to $u = -1$. Since $|u, \phi\rangle = d_+^{1/2}(u)|0\rangle + e^{i\phi} d_-^{1/2}(u)|1\rangle$, with $d_\pm(u) \equiv \frac{1}{2}(1 \pm u)$ any $|\mathbf{u}, \mathbf{\Phi}\rangle$ may be decomposed as

$$|\mathbf{u}, \mathbf{\Phi}\rangle = \sum_A \prod_{i\in A} d_+^{1/2}(u_i) \prod_{j \in A^c} d_-^{1/2}(u_j) \exp\left(i \sum_{j\in A^c} \phi_j\right) |A\rangle \tag{13}$$

where the index A runs over all the subsets of $\{1, 2, \ldots, n\}$, A^c is the complement of A, and $|A\rangle$ is the basis state corresponding to the binary sequence with zeros in the positions indexed by members of A. It is easy to see that the off diagonal elements of the density matrix

$$\rho = \int d\mathbf{u} P(\mathbf{u}) \int d\mathbf{\Phi} |\mathbf{u}, \mathbf{\Phi}\rangle\langle\mathbf{u}, \mathbf{\Phi}| \tag{14}$$

vanish in this basis, thus reducing the von Neumann entropy to a Shannon entropy. The new problem becomes: find the distibution $P(\mathbf{u})$ which minimizes

$$\int d\mathbf{u}P(\mathbf{u})\log P(\mathbf{u}) + \lambda\sum_A \overline{d}_A \log \overline{d}_A - \mu\int d\mathbf{u}P(\mathbf{u}) \tag{15}$$

with $\overline{d}_A = \int d\mathbf{u}P(\mathbf{u})d_A(\mathbf{u})$ and $d_A(\mathbf{u}) = \prod_{i\in A} d_+(u_i)\prod_{j\in A^c} d_-(u_j)$. Setting the functional derivative with respect to $P(\mathbf{u})$ to zero and using the normalization condition to eliminate μ yields the following system of equations

$$P(\mathbf{u}) = Z^{-1}\exp(\lambda\sum_A c_A d_A(\mathbf{u})) \tag{16}$$

$$\log \overline{d}_A = c_A \tag{17}$$

where

$$Z = \int d\mathbf{u}\exp(\lambda\sum_A c_A d_A(\mathbf{u})) \tag{18}$$

We shall now prove that the solution to the above equations is remarkably one that factorizes into identical single qubit distributions. We guess a solution of the form

$$P(\mathbf{u}) = z^{-n}\prod_i \exp(\lambda_+ d_+(u_i) + \lambda_- d_-(u_i)) \tag{19}$$

where $z = \int du\, e^{\lambda_+ d_+(u)+\lambda_- d_-(u)}$. Then comparing with (16) we find that $c_A = |A|\lambda_+ + (1-|A|)\lambda_-$. On the other hand

$$\log \overline{d}_A = |A|\log \overline{d}_+ + (1-|A|)\log \overline{d}_- \tag{20}$$

where

$$\overline{d}_\pm \equiv z^{-1}\int du\, d_\pm(u) e^{\lambda_+ d_+(u)+\lambda_- d_-(u)} = \lambda_\pm \tag{21}$$

Hence (16) and (17) are simultaneously satisfied. Since the expression (15) is convex in $P(\mathbf{u})$ (because $\lambda \geq 0$), the solution we have found is indeed a minimum, and it is unique.

Therefore $R_n(S) = R_1(S)$ and a simple calculation yields the following parametrization

$$R_1(\lambda) = \frac{\lambda}{e^\lambda - 1} - 1 + \log\left(\frac{\lambda e^\lambda}{e^\lambda - 1}\right) \tag{22}$$

$$S(\lambda) = h_2\left(\frac{1}{\lambda} - \frac{1}{e^\lambda - 1}\right) \tag{23}$$

where the $\lambda \in (0, \infty)$ and $h_2(p) = -p\log p - (1-p)\log(1-p)$ is the binary Shannon entropy function. The single letter conditional distribution that achieves this curve is given by

$$Q^\lambda(y|x) = P^\lambda(x|y) = \frac{1}{4\pi}\frac{\lambda}{e^\lambda - 1}e^{\lambda|\langle x|y\rangle|^2} \tag{24}$$

The curve is readily found to be convex. We shall now establish $R(S) = R_1(S)$. First we show that $R_1(S)$ is a lower bound on (R, S) pairs attained with codes B_n of size K and blocklength n. Indeed

$$R = \frac{1}{n}\log_2 K \geq \frac{1}{n}H(\mathbf{Y}) \geq \frac{1}{n}I(\mathbf{X};\mathbf{Y}) \geq R_n(\overline{S}(B_n)) = R_1(\overline{S}(B_n)) \tag{25}$$

The first two inequalities are elementary results of classical information theory [7] and the third one is a consequence of the definition (3). Hence $R_1(S)$ is a lower bound on $R(S)$. To complete the proof we have to show that this bound is attainable. In the following we fix λ and drop the superscript in (24). We will construct a source code that will come arbitrarily close to the point $(R_1(\lambda), S(\lambda))$. The idea is to simulate the noisy single letter channel defined by $Q(y|x)$ by the average effect a deterministic map involving large strings of letters has on the ith letter.

We pick a set $\widehat{\mathcal{X}}$ of points in $\mathcal{X}$ and construct a tiling of $\mathcal{X}$, $\mathcal{T} = \{\mathcal{T}(\widehat{x})|\widehat{x} \in \widehat{\mathcal{X}}\}$, where each $\mathcal{T}(\widehat{x})$ is a region containig $\widehat{x}$ such that the $\mathcal{T}(\widehat{x})$ are disjoint for different $\widehat{x}$ and their union is $\mathcal{X}$. For a small enough ϵ we may choose $\mathcal{T}$ so that the area $A_{\mathcal{T}(\widehat{x})} = \int_{\mathcal{T}(\widehat{x})} dx$ covered by each $\mathcal{T}(\widehat{x})$ is greater than ϵ^2 and the maximum distance of any point of $\mathcal{T}(\widehat{x})$ from $\widehat{x}$ is less than ϵ. Now we define the discretized distributions $\widehat{q}(\widehat{y}) \equiv \int_{\mathcal{T}(\widehat{y})} dy q(y)$ and $\widehat{P}(\widehat{x}|\widehat{y}) \equiv \int_{\mathcal{T}(\widehat{x})} dx P(x|\widehat{y})$ for $\widehat{x}, \widehat{y} \in \widehat{\mathcal{X}}$. From these we further define the marginal distribution $\widehat{p}(\widehat{x}) \equiv \sum_{\widehat{y}\in\widehat{\mathcal{X}}} \widehat{P}(\widehat{x}|\widehat{y})\widehat{q}(\widehat{y})$. These are now probabilities not probability densities. We also define the conditional probability density of a related continuous distribution defined on $\mathcal{X}$: $P^d(x|\widehat{y}) \equiv \widehat{P}(\widehat{x}|\widehat{y})/A_{\mathcal{T}(\widehat{x})}$ for $x \in \mathcal{T}(\widehat{x})$. Since the gradient of $P(x|y)$ is bounded, we see that $|P^d(x|\widehat{y}) - P(x|\widehat{y})| < c\epsilon$ for some constant c. The mutual information of the discretized distributions is given by

$$\widehat{I} = \sum_{\widehat{x}\in\widehat{\mathcal{X}}}\sum_{\widehat{y}\in\widehat{\mathcal{X}}} \widehat{P}(\widehat{x}|\widehat{y})\widehat{q}(\widehat{y}) \log \frac{\widehat{P}(\widehat{x}|\widehat{y})}{\widehat{p}(\widehat{x})} \tag{26}$$

It is easy to see that $|\widehat{I} - R_1(\lambda)| < c'\epsilon$ for some constant c' since the gradient of $P(x|y)$ and $\mathcal{X}$ itself are both bounded, and $P(x|y) > 0 \;\forall x, y \in \mathcal{X}$.

We now follow the joint typicality source coding scheme [7], Section 13.6 based on the discretized distributions. We define a *δ-typical sequence* $\widehat{\mathbf{x}} \in \widehat{\mathcal{X}}^n$ with respect to the distribution $\widehat{p}(\widehat{x})$ as one that satisfies

$$\left|\frac{N(\widehat{a}|\widehat{\mathbf{x}})}{n} - \widehat{p}(\widehat{a})\right| < \frac{\delta}{|\widehat{\mathcal{X}}|} \tag{27}$$

where $N(\widehat{a}|\widehat{\mathbf{x}})$ is the number of occurences of $\widehat{a} \in \widehat{\mathcal{X}}$ in the sequence $\widehat{\mathbf{x}}$. We call the set of all such typical sequences the *typical set* $T_\delta(\widehat{p})$. One similarly defines the *jointly typical set* $T_\delta(\widehat{P}\widehat{q})$ of *pairs* of typical sequences $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) \in (\widehat{\mathcal{X}} \times \widehat{\mathcal{X}})^n$ with respect to the distribution $\widehat{P}(\widehat{x}|\widehat{y})\widehat{q}(\widehat{y})$ [7]. Note that the cardinality of $\widehat{\mathcal{X}}$ is bounded as $4/\epsilon^2 < |\widehat{\mathcal{X}}| < 4\pi/\epsilon^2$. We now generate the source code $B_n = \{\widehat{\mathbf{y}}_1, \widehat{\mathbf{y}}_2, \ldots, \widehat{\mathbf{y}}_K\}$ of rate $R = n^{-1}\log_2 K$ by choosing each $\widehat{\mathbf{y}}_k$ randomly from the typical set $T_\delta(\widehat{q})$. A given $\mathbf{x} \in \mathcal{X}^n$ gets mapped into $\widehat{\mathbf{y}} \in B_n$ such that $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}})$ is jointly typical, where $\widehat{x} \in \widehat{\mathcal{X}}$ is such that $x_i \in \mathcal{T}(\widehat{x}_i), \forall i$. If there is more than one such $\widehat{\mathbf{y}}$ it is mapped into the first one in lexicographical order. If there is no such $\widehat{\mathbf{y}}$ it gets encoded as an error message.

Following [7] we see that given $\widehat{\mathbf{x}}$ there can be two kinds of errors: either $\widehat{\mathbf{x}}$ is not typical at all, or it is typical but not jointly typical with any $\widehat{\mathbf{y}} \in B_n$. They show that the first kind of error occurs with probability $p_1 < \epsilon$ for large enough n. The probability p_2 of the second kind of error is bounded by

$$p_2 \leq e^{-2^{n(R-\widehat{I}-\epsilon_1)}} \tag{28}$$

where $\epsilon_1 \to 0$ as $\delta \to 0$ and $n \to \infty$. So we may choose a sufficiently small $\delta < 2\epsilon$ and sufficiently large n so that $\epsilon_1 < \epsilon$. Then for $R > R_1(\lambda) + (c'+1)\epsilon > \widehat{I} + \epsilon_1$ and sufficiently large n we can make $p_2 < \epsilon$ also, thus bounding the total error probability by 2ϵ.

In the case of successful encoding $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}})$ is jointly typical. Viewing the elements of $\widehat{\mathbf{x}}$ and $\widehat{\mathbf{y}}$ as random variables $\widehat{X}_i$ and $\widehat{Y}_i$ we have

$$Pr(\widehat{X}_i = \widehat{x}_i|\widehat{Y} = \widehat{y}_i) = \frac{Pr(\widehat{X}_i = \widehat{x}_i, \widehat{Y}_i = \widehat{y}_i)}{Pr(\widehat{Y}_i = \widehat{y}_i)} \in \left(\widehat{P}(\widehat{x}_i|\widehat{y}_i) - \frac{2\delta}{|\widehat{\mathcal{X}}|}, \widehat{P}(\widehat{x}_i|\widehat{y}_i) + \frac{2\delta}{|\widehat{\mathcal{X}}|}\right) \tag{29}$$

for small enough δ, which easiliy follows from $\widehat{y} \in T_\delta(\widehat{q})$ and $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}}) \in T_\delta(\widehat{P}\widehat{q})$. Since the conditional distribution of X_i given $\widehat{X}_i$ is uniform and $X_i \to \widehat{X}_i \to \widehat{Y}_i$ form a Markov chain, we see that the conditional distibution $\widetilde{P}(x|\widehat{y})$ of the X_i probability density given $\widehat{Y}_i$ is related to $P^d(x|\widehat{y})$:

$$|\widetilde{P}(x|\widehat{y}) - P^d(x|\widehat{y})| < \frac{2\delta}{\epsilon^2|\widehat{\mathcal{X}}|} < \frac{\delta}{2} \tag{30}$$

and so by the triangle inequality $|\widetilde{P}(x|\widehat{y}) - P(x|\widehat{y})| < \delta/2 + c\epsilon < (c+1)\epsilon$. Now define $\mathbf{Z} \equiv \mathbf{\Omega}_{\widehat{\mathbf{Y}} \to \mathbf{0}}\mathbf{X}$. In words, $\mathbf{Z}$ is the random variable defined by acting on $\mathbf{X}$ with the element of $\mathcal{R}^n$ that takes $\widehat{\mathbf{Y}}$ into $\mathbf{0}$. We have the following string of inequalities

$$\overline{S}(B_n) = \frac{1}{n} E_{\widehat{\mathbf{Y}}} S(E_{\mathbf{X}|\widehat{\mathbf{Y}}} |\mathbf{X}\rangle\langle\mathbf{X}|) = \frac{1}{n} E_{\widehat{\mathbf{Y}}} S(\rho_{\widehat{\mathbf{Y}}}) \leq \frac{1}{n} E_{\widehat{\mathbf{Y}}} \sum_i S(\rho^{(i)}_{\widehat{\mathbf{Y}}}) \leq S(\overline{\rho}) \tag{31}$$

where $\rho_{\widehat{\mathbf{Y}}} \equiv E_{\mathbf{X}|\widehat{\mathbf{Y}}} |\mathbf{Z}\rangle\langle\mathbf{Z}|$, $\rho^{(i)}_{\widehat{\mathbf{Y}}}$ is the restriction of $\rho_{\widehat{\mathbf{Y}}}$ to the ith letter, and $\overline{\rho} \equiv \frac{1}{n} E_{\widehat{\mathbf{Y}}} \sum_i \rho^{(i)}_{\widehat{\mathbf{Y}}}$. The two inequalities are consequences of subadditivity and concavity of the von Neumann entropy respectively. First note that the conditional probability density $\widetilde{P}(z|\widehat{y})$ of Z_i given $\widehat{Y}_i$ satisfies $|\widetilde{P}(z|\widehat{y}) - P(x|\widehat{y})| < (c+1)\epsilon$ independently of i and $\widehat{\mathbf{y}} \in B_n$. Consequently both $\rho^{(i)}_{\widehat{\mathbf{y}}}$ and $\overline{\rho}$ have an entry-wise error of at most $4\pi(c+1)\epsilon$ relative to the optimal density matrix $\rho(\lambda) \equiv \int dx P(x|0)|x\rangle\langle x|$. Finally, $\overline{S}(B_n) \leq S(\overline{\rho}) < S(\lambda) + 4\pi(c+1)k\epsilon$ where $k = h_2'(\frac{1}{\lambda} - \frac{1}{e^\lambda - 1})$.

In conlcusion, for any $R > R_1$ we can find a sufficiently small ϵ so that $R > R_1 + (c'+1)\epsilon$, and for a large enough n we can construct a code B_n of rate R that achieves $S < 4\pi(c+1)k\epsilon$ with probability greater than $1 - 2\epsilon$.

The RSP protocol is as follows: Alice wishes to remotely prepare a string of n qubits using an (R, S) source code. She identifies the corresponding codeword and rotates the original string by the map that sends the codeword to $\mathbf{0}$, and prepares these qubits in her laboratory. She may Schumacher compress them without additional blocking to Sn qubits, treating them as i.i.d. with density matrix $\overline{\rho}$. She teleports these to Bob using $2Sn$ classical bits and Sn ebits. A further Rn bits are sent in order to convey the codeword. Bob simply reverses Alice's steps in his laboratory, thus recovering an asymptotically faithful copy of the qubits to be prepared. The corresponding point in the ebits vs. bits plane is $(S, R + 2S)$. The optimal ebits vs. bits tradeoff is shown in Fig.1.

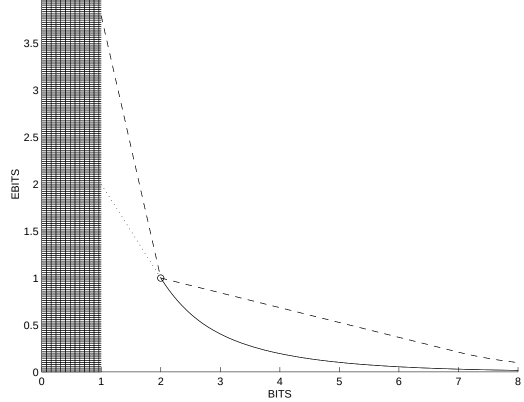


FIG. 1 Ebits v. bits for remote state preparation. The solid curve represents our exact curve and the dotted line our lower bound. The dashed curve is the upper bound obtained in [1]. The shaded region is forbidden by causality.

Incidentally, this low-entanglement result gives rise to a lower bound for the high-entanglement region as shown by the dotted line of unit slope, tangent to the low-entanglement curve at $(2, 1)$. Indeed, if it were possible to achieve a point below this bound then by time sharing with another point on the convex low-entanglement curve one would be able to improve on the $(2, 1)$ point corresponding to teleportation, thus violating causality as already noted. Future efforts should be directed at narrowing down the high-entanglement curve which we now know to lie somewhere between our lower bound and the upper bound found in [1].

We are grateful to N.D.Mermin for bringing reference [1] to our attention. This research was supported in part by the ARO-administered MURI Grant No. DAAG-19-99-1-0125 and NSF Grant CCR-9980616.

References

[1] C.H.Bennett, D.P.DiVincenzo, J.A.Smolin, B.M.Terhal and W.K.Wooters, quant-ph/0006044 (2000)

[2] C.H.Bennett, G.Brassard, C.Crépeau, R.Josza, A.Peres and W.K.Wooters, *Phys.Rev.Lett.* **70**, 1895 (1993)

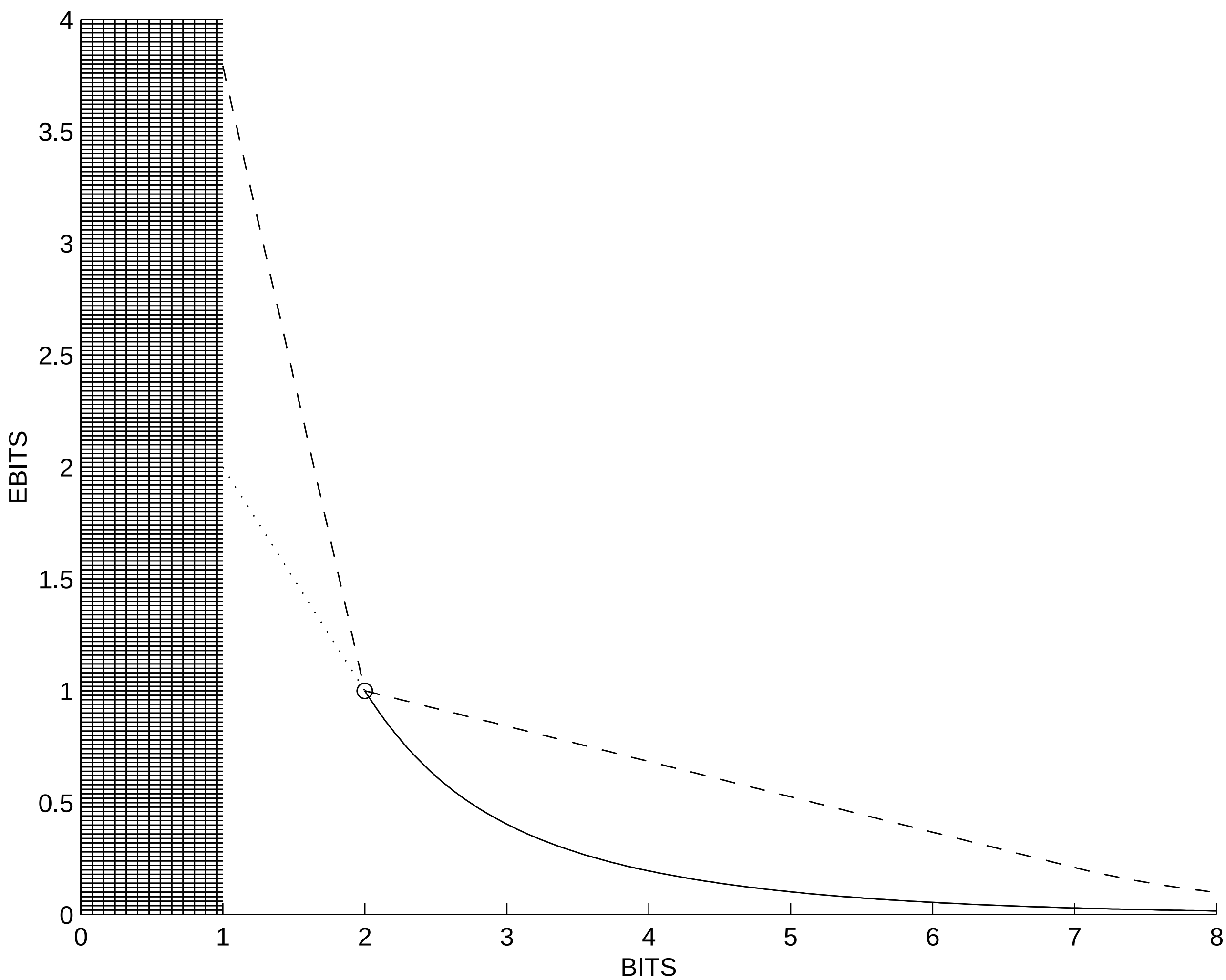
[3] H-K Lo, quant-ph/9912009 (1999)

[4] C.H.Bennett and S.J.Wiesner, *Phys.Rev.Lett.* **69**, 2881 (1992)

[5] B.Schumacher, *Phys.Rev.A* **51**, 2738 (1995); R.Josza and B.Schumacher, *J.Mod.Opt* **41**, 2343 (1994)

[6] T.Berger, *Rate distortion theory, Prentice Hall* (1971)

[7] T.Cover and J.Thomas, *Elements of information theory, Wiley and Sons* (1991)



4
3.5
3
2.5
2
1.5
1
0.5
0
EBITS
0
1
2
3
4
5
6
7
8
BITS